Module 5

- **Std. deviat^n**

  Std. deviat^n can never be -ve.

  - **Populat^n**

    $$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N}} = \frac{\sqrt{N\Sigma x^2 - (\Sigma x)^2}}{N}$$

  - **Sample**

    $$S = \sqrt{\frac{\Sigma(x - \bar{x})^2}{N - 1}}$$

- **Correlat^n Coefficient (Denoted by *r*) ↓**

  • Also called Correlat^n Index.
  • Its value lies between -1 & 1.
  • Value of Perfect Correlat^n is either -1 or 1.
  • Value of Imperfect Correlat^n lies between -1 & 1.

- **Karl-Pears^n's model of correlat^n coefficient**

$$r = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2 \Sigma(Y - \bar{Y})^2}}$$

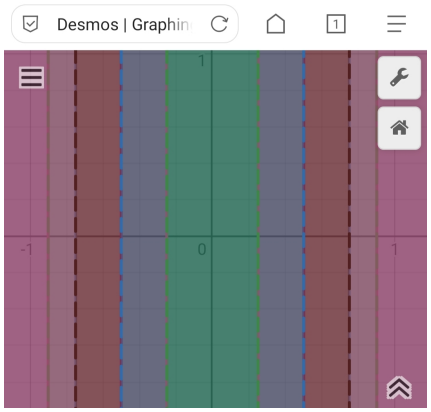
$$r = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

  - **N.B.**

$$Cov(X, Y) = \frac{\Sigma xy}{N} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{N} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{N^2}$$

  - **Some facts on *r***

Value of r	Correlation Degree
$r = 1$	Perfect +ve
$r = -1$	Perfect -ve
$r \in (0, 1)$	Imperfect +ve
$r \in (-1, 0)$	Imperfect -ve
$r = 0$	No Correlation
$r \in (-\infty, -1] \cup [1, \infty)$	Erroneous Result
$r \in (-\infty, -0.9] \cup [0.9, \infty)$	Very High Degree
$r \in (-0.9, -0.75] \cup [0.75, 0.9)$	Fairly High Degree
$r \in (-0.75, -0.5] \cup [0.5, 0.75)$	Moderate Degree
$r \in (-0.5, -0.25] \cup [0.25, 0.5)$	Low Degree
$r \in (-0.25, 0.25)$	Very Low Degree



- Spearman's rank correlat^{n} coefficient

$$r_s = 1 - \frac{6\Sigma D^2}{N^3 - N}$$

$$\Sigma D^2 = \Sigma (R_X - R_Y)^2$$

- Regress^{n} lines
  - Least sq. method

    Also called Regress^{n} Equn. using Normal Equns.

    - Y on X

$$Y = a + bX$$

$$\Sigma Y = aN + b\Sigma X$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

    - X on Y

$$X = c + dY$$

$$\Sigma X = cN + d\Sigma Y$$

$$\Sigma XY = c\Sigma Y + d\Sigma Y^2$$

- Deviat^{n} method from actual mean
  - Y on X

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = \frac{N\Sigma XY - \Sigma X\Sigma Y}{N\Sigma X^2 - (\Sigma X)^2}$$

  - X on Y

$$X - \bar{X} = b_{xy}(Y - \bar{Y})$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = \frac{N\Sigma XY - \Sigma X\Sigma Y}{N\Sigma Y^2 - (\Sigma Y)^2}$$

  - N.B.

$$r = \sqrt{b_{xy} \cdot b_{yx}}$$

- N.B.

$$Y = a + bX$$

$$a = \bar{Y} - b_{yx}\bar{X}$$

$$b = b_{yx}$$

$$X = c + dY$$

$$c = \bar{X} - b_{xy}\bar{Y}$$

$$d = b_{xy}$$

- Statistical Inference ↓
- Pt. Estimatⁿ
  - MLE Method

$$Step - 1 : \; L = \prod_{i=1}^{n} f(x_n, \theta)$$

$$Step - 2 : \; ln(L) = ln\left(\prod_{i=1}^{n} f(x_n, \theta) \right)$$

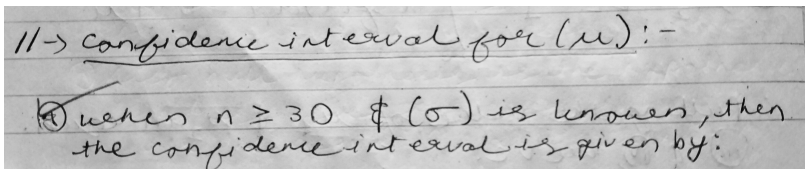
$$Step - 3 : \; \frac{\partial}{\partial \theta}(ln(L)) = 0$$

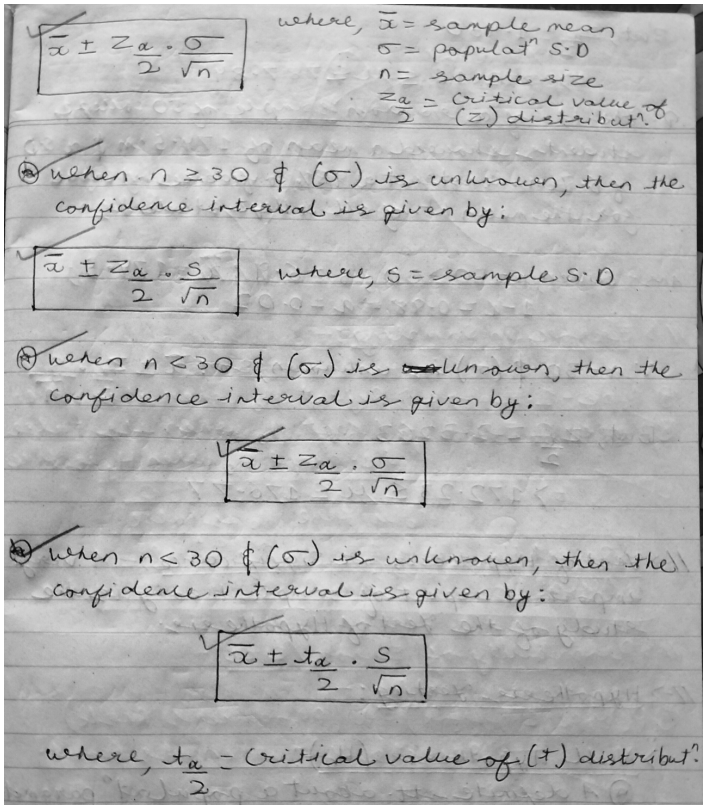
Uses populatⁿ mean everytime.

  - MOM Method
- Interval Estimatⁿ

$$CONFIDENCE\ INTERVAL = 1 - \alpha$$
$$SIGNIFICANCE\ LEVEL = \alpha$$

  - Confidence Interval for μ





$$\bar{x} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

where, $\bar{x}$ = sample mean
σ = populat^n S.D
n = sample size
$Z_{\frac{\alpha}{2}}$ = Critical value of (Z) distribut^n.

* When $n \geq 30$ & (σ) is unknown, then the confidence interval is given by:

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$ where, s = sample S.D

* When $n < 30$ & (σ) is unknown, then the confidence interval is given by:

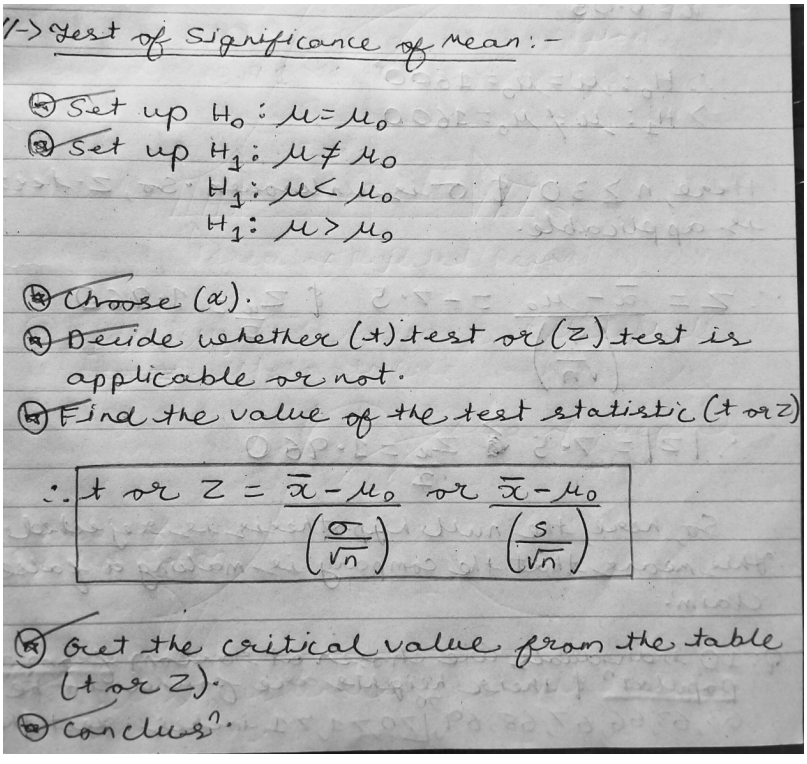
$$\bar{x} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

* When $n < 30$ & (σ) is unknown, then the confidence interval is given by:

$$\bar{x} \pm t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

where, $t_{\frac{\alpha}{2}}$ = Critical value of (t) distribut^n.

- **Hypothesis Test**

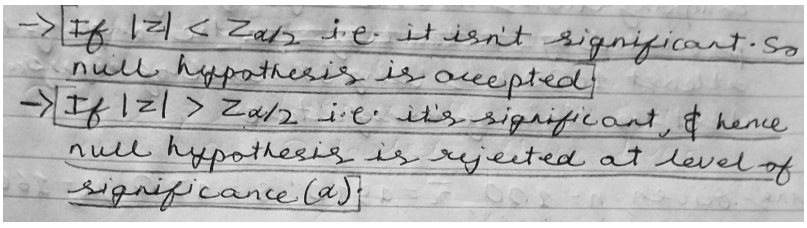


-> Test of Significance of mean:-

* Set up $H_0 : \mu = \mu_0$
* Set up $H_1 : \mu \neq \mu_0$
  $H_1 : \mu < \mu_0$
  $H_1 : \mu > \mu_0$
* Choose (α).
* Decide whether (t) test or (Z) test is applicable or not.
* Find the value of the test statistic (t or Z)

$$\therefore t \text{ or } Z = \frac{\bar{x} - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} \text{ or } \frac{\bar{x} - \mu_0}{\left(\frac{s}{\sqrt{n}}\right)}$$

* Get the critical value from the table (t or Z).
* Conclusn.



-> If $|Z| < Z_{\alpha/2}$ i.e. it isn't significant. So null hypothesis is accepted.
-> If $|Z| > Z_{\alpha/2}$ i.e. it's significant, & hence null hypothesis is rejected at level of significance (α).

- **N.B.**

// → Error in Sampling :-

* Type-1 Error → Reject H_0 when it's True.
* Type-2 Error → Accept H_0 when it's False
  i.e. accept H_0 when H_1 is True.

	H_0 is True	H_0 is False
Reject H_0	Type-1 Error	Correct decision
Accept H_0	Correct decision	Type-2 Error

// → Probability of Error / Size :-

* P(Type-1 Error) = α (Producer's Risk)
  → Reject a lot when it's Good.

* P(Type-2 Error) = β (Consumer's Risk)
  → Accept a lot when it's Bad.

Here, (α) is called Level of significance or Area of rejected region.